Exact lower bounds on the exponential moments of Winsorized and truncated random variables

Iosif Pinelis

Michigan Technological University

Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@mtu.edu

Abstract: Exact lower bounds on the exponential moments of $\min(y, X)$ and $X\,\mathrm{I}\{X < y\}$ are provided given the first two moments of a random variable X. These bounds are useful in work on large deviations probabilities and nonuniform Berry-Esseen bounds, when the Cramér tilt transform may be employed. Asymptotic properties of these lower bounds are presented. Comparative advantages of the Winsorization $\min(y, X)$ over the truncation $X\,\mathrm{I}\{X < y\}$ are demonstrated.

AMS 2000 subject classifications: Primary 60E15; secondary 60E10, 60F10, 60F05.
Keywords and phrases: exponential moments, exact lower bounds, Winsorization, truncation, large deviations, nonuniform Berry-Esseen bounds, Cramér tilt transform.

1. Introduction

Cramér's tilt transform of a random variable (r.v.) X is a r.v. X_c such that

$$\mathsf{E}f(X_c) = \frac{\mathsf{E}f(X)e^{c\,X}}{\mathsf{E}e^{c\,X}} \tag{1.1}$$

for all nonnegative Borel functions f, where c is a real parameter. This transform is an important tool in the theory of large deviation probabilities $\mathsf{P}(X > x)$, where $x > 0$ is a large number; then the appropriate value of the parameter c is positive. As e.g. in the proof of [16, Theorem 2.3], one often needs to bound from above the f-moment $\mathsf{E}f(X_c)$ of the c-tilted r.v. X_c for a nonnegative f – and therefore one needs to bound the denominator $\mathsf{E}e^{c\,X}$ in (1.1) from below. If $\mathsf{E}X = 0$, this can be done quite easily: by Jensen's inequality, $\mathsf{E}e^{c\,X} \geqslant 1$.

A usual problem with this approach occurs when the right tail of X is too heavy for $\mathsf{E}e^{cX}$ to be finite and hence for the transform to make sense. The standard cure in such situations is to truncate the r.v. X, say to $T_y(X) := X\,\mathrm{I}\{X \leqslant y\}$ for some real number $y > 0$, where $\mathrm{I}\{\cdot\}$ is the indicator function. Then, of course, $\mathsf{E}e^{c\,T_y(X)} < \infty$ for any $c > 0$. However, now instead of the condition $\mathsf{E}X = 0$ one has $\mathsf{E}T_y(X) \leqslant 0$, and the inequality $\mathsf{E}e^{c\,T_y(X)} \geqslant 1$ (in

place of $\mathsf{E}e^{c\,X} \geqslant 1$) will not hold in general. In fact, $\mathsf{E}e^{c\,T_y(X)}$ can be however small for some $c > 0$, even if one imposes a restriction such as $\mathsf{E}X^2 \leqslant \sigma^2$ for a given real $\sigma > 0$ – see the discussion in Subsection 2.3.

A much better way to cut off the right tail of the distribution of X is the so-called Winsorization. That is, instead of the truncation $T_y(X)$, one deals with $W_y(X) := y \wedge X = \min(y, X)$. Clearly, $W_y(X) \geqslant T_y(X)$ and hence $\mathsf{E}e^{c\,W_y(X)} \geqslant \mathsf{E}e^{c\,T_y(X)}$ for $c > 0$. Moreover, it turns out that for any given real $\sigma > 0$ and $y > 0$ the infimum of $\mathsf{E}e^{c\,W_y(X)}$ over all $c > 0$ and all r.v.'s X with $\mathsf{E}X \geqslant 0$ and $\mathsf{E}X^2 \leqslant \sigma^2$ is strictly positive; furthermore, it decreases slowly from 1 to 0 as σ increases from 0 to ∞. These properties of Winsorization make it a clear winner over truncation in many relevant situations.

2. Results

Take any real $\sigma > 0$. Let X denote any r.v. with

$$\mathsf{E}X \geqslant 0 \quad \text{and} \quad \mathsf{E}X^2 \leqslant \sigma^2.$$

For any positive real a and b, let $X_{a,b}$ stand for any zero-mean r.v. with values in the two-point set $\{-a, b\}$; thus, the distribution of $X_{a,b}$ is uniquely determined by a and b. Note also that $\mathsf{E}X_{a,b}^2 = ab$.

2.1. Winsorization

Consider the Winsorization function defined by the formula

$$W(x) := 1 \wedge x. \tag{2.1}$$

The following proposition allows one to define the terms in which to express the exact lower bounds on $\mathsf{E}e^{c\,W(X)}$.

Proposition 2.1. *Take any real $c > 0$.*

(I) For any real a, let

$$b^*_{a,c} := \tfrac{2(e^{c+ac}-1)-ac}{c}. \tag{2.2}$$

Then the equation $a\,b^*_{a,c} = \sigma^2$ has a unique positive root, say $a_{c,\sigma}$, so that

$$\{a_{c,\sigma}\} = \big\{a > 0\colon a\,b^*_{a,c} = \sigma^2\big\}. \tag{2.3}$$

(II) The expression

$$\ell_1(a) := \ell_1(a,\sigma) := \ln\tfrac{a}{\sigma^2} - \tfrac{2(a+1)(a-\sigma^2)}{a^2+\sigma^2} \tag{2.4}$$

switches in sign exactly once, from $-$ to $+$, as a increases from 0 to σ^2. Therefore, one can uniquely define a_σ by the formula

$$\{a_\sigma\} = \big\{a \in (0,\sigma^2)\colon \ell_1(a) = 0\big\}. \tag{2.5}$$

The proofs are deferred to Section 3.

Now we are ready to define three more symbols:

$$b_{c,\sigma} := \sigma^2/a_{c,\sigma}; \tag{2.6}$$

$$b_\sigma := \sigma^2/a_\sigma; \tag{2.7}$$

$$c_\sigma := \tfrac{\ln b_\sigma}{1+a_\sigma}. \tag{2.8}$$

Theorem 2.2. *For any real $c > 0$*

$$\begin{aligned}\mathsf{E}\exp\{c\,W(X)\} \geqslant L_{W;c,\sigma} &:= \mathsf{E}\exp\{c\,W(X_{a_{c,\sigma},b_{c,\sigma}})\} &(2.9)\\ \geqslant L_{W;\sigma} &:= \mathsf{E}\exp\big\{c_\sigma\,W(X_{a_\sigma,b_\sigma})\big\}. &(2.10)\end{aligned}$$

Moreover, inequality (2.9) *is strict unless $X \overset{\mathrm{D}}{=} X_{a_{c,\sigma},b_{c,\sigma}}$, where $\overset{\mathrm{D}}{=}$ denotes the equality in distribution, and inequality* (2.10) *is strict unless $c = c_\sigma$. Furthermore,*

$$a_\sigma = a_{c_\sigma,\sigma} \quad \textit{and} \quad b_\sigma = b_{c_\sigma,\sigma} = b^*_{a_\sigma,c_\sigma}, \tag{2.11}$$

so that (2.10) *turns into equality if and only if $c = c_\sigma$.*

In addition to being zero-mean, each of the r.v.'s $X_{a_{c,\sigma},b_{c,\sigma}}$ and X_{a_σ,b_σ} has variance σ^2, in view of (2.6) and (2.7). Moreover, by (2.5) and (2.7), $b_\sigma > 1$ and hence $c_\sigma > 0$ by (2.8). Thus, (2.9) provides an exact lower bound on $\mathsf{E}\exp\{c\,W(X)\}$ for a fixed $c > 0$, while (2.10) provides an exact lower bound on $\mathsf{E}\exp\{c\,W(X)\}$ over all $c > 0$.

Let us now describe the asymptotics of the bounds $L_{W;c,\sigma}$ and $L_{W;\sigma}$ for $\sigma \downarrow 0$ and $\sigma \to \infty$. As usual, we write $a \sim b$ if $\frac{a}{b} \to 1$.

Proposition 2.3.

(I) For any real $c > 0$

$$\begin{aligned}L_{W;c,\sigma} - 1 &\sim \tfrac{-c^2}{4(e^c-1)}\,\sigma^2 \quad \textit{as } \sigma \downarrow 0, &(2.12)\\ L_{W;c,\sigma} &\sim \tfrac{4e^c}{c^2}\,\tfrac{\ln^2\sigma}{\sigma^2} \quad \textit{as } \sigma \to \infty. &(2.13)\end{aligned}$$

(II) The expression

$$f(t) := \ln t + 2(1-t) \tag{2.14}$$

switches in sign exactly once, from $-$ to $+$, as t increases from 0 to 1; Therefore, one can uniquely define t_* by the formula

$$\{t_*\} = \big\{t \in (0,1)\colon f(t) = 0\big\}; \tag{2.15}$$

in fact, $t_* = 0.203\ldots$.

(III)

$$\begin{aligned}L_{W;\sigma} - 1 &\sim -(1-t_*)t_*\,\sigma^2 \quad \textit{as } \sigma \downarrow 0, &(2.16)\\ L_{W;\sigma} &\sim e^2\,\tfrac{\ln^2\sigma}{\sigma^2} \quad \textit{as } \sigma \to \infty, &(2.17)\end{aligned}$$

(IV) Comparing (2.12) *with* (2.16), *and* (2.13) *with* (2.17)*:*

$$\inf_{c>0} \tfrac{-c^2}{4(e^c-1)} = -(1-t_*)t_* \text{ – attained at } c = -\ln t_* = 1.593\ldots; \tag{2.18}$$

$$\inf_{c>0} \tfrac{4e^c}{c^2} = e^2 \qquad \text{ – attained at } c = 2. \tag{2.19}$$

Thus, the asymptotic expression in (2.16) *is the minimum in* $c > 0$ *of that in* (2.12)*, and the asymptotic expression in* (2.17) *is the minimum in* $c > 0$ *of that in* (2.13)*.*

Note that the convergence for $\sigma \to \infty$ in Proposition 2.3 is very slow. E.g., the ratio $e^2 \frac{\ln^2\sigma}{\sigma^2}/L_{W;\sigma}$ of the terms in (2.17) is $1.201\ldots$ for σ as large as 10^{10}.

The relations $a_\sigma \sim t_*\sigma^2$ as $\sigma \downarrow 0$ and $a_\sigma \sim \frac12\ln(\sigma^2)$ as $\sigma \to \infty$, to be established in the proof of part (III) of Proposition 2.3, suggest, and numerical calculations confirm, that a good initial approximation for solving the equation $\ell_1(a) = 0$ in (2.5) for a_σ is $a = \frac12\ln(1+2t_*\sigma^2) \approx \frac12\ln(1+0.406\,\sigma^2)$.

2.2. Truncation

Consider the truncation function T defined by the formula

$$T(x) := x\,\mathrm{I}\{x<1\}. \tag{2.20}$$

The following proposition allows one to define the terms in which to express the exact lower bounds on $\mathsf{E}e^{cT(X)}$.

Proposition 2.4. *Take any real* $c > 0$*. For any real* a*, let*

$$B^*_{a,c} := \tfrac{2(e^{ac}-1)-ac}{c};$$

cf. (2.2)*. Then one can uniquely define* A_σ *and* $A_{c,\sigma}$ *by the formulas*

$$\{A_c\} = \big\{a>0\colon B^*_{a,c} = 1\big\}, \tag{2.21}$$

$$\{A_{c,\sigma}\} = \big\{a>0\colon a\,B^*_{a,c} = \sigma^2\big\}, \tag{2.22}$$

because each of the equations on the right-hand sides of (2.21) *and* (2.22) *has a unique root* $a > 0$*. Moreover, one has the implication*

$$A_c \leqslant \sigma^2 \implies A_{c,\sigma} \geqslant A_c. \tag{2.23}$$

We shall need one more definition:

$$B_{c,\sigma} := \sigma^2/A_{c,\sigma}.$$

Theorem 2.5. *For any real* $c > 0$

$$\mathsf{E}\exp\{cT(X)\} \geqslant L_{T;c,\sigma} := \begin{cases} \mathsf{E}\exp\{cT\big(X_{\sigma^2,1}\big)\} & \text{if } \sigma^2 \leqslant A_c, \\ \mathsf{E}\exp\{cT\big(X_{A_{c,\sigma},B_{c,\sigma}}\big)\} & \text{if } \sigma^2 \geqslant A_c. \end{cases} \tag{2.24}$$

Moreover, inequality (2.24) *is strict unless* X *equals* $X_{\sigma^2,1}$ *or* $X_{A_{c,\sigma},B_{c,\sigma}}$ *in distribution, depending on whether* $\sigma^2 \leqslant A_c$ *or* $\sigma^2 \geqslant A_c$*.*

To complete this subsection, let us describe the asymptotics of the bound $L_{T;c,\sigma}$ for $\sigma \downarrow 0$ and $\sigma \to \infty$ – cf. Proposition 2.3.

Proposition 2.6. *For any real $c > 0$*

$$L_{T;c,\sigma} - 1 \sim -c\,\sigma^2 \quad \text{as } \sigma \downarrow 0, \tag{2.25}$$

$$L_{T;c,\sigma} \sim \tfrac{4}{c^2}\,\tfrac{\ln^2\sigma}{\sigma^2} \ \text{ as } \sigma \to \infty. \tag{2.26}$$

2.3. Winsorization and truncation: discussion and comparison

The Winsorization function W and the truncation function T as defined by (2.1) and (2.20) "cut" a given value x at the level 1. However, by simple rescaling it is easy to restate the results for any positive "cut" level y. Indeed, one may consider $W_y(x) := y \wedge x$ and $T_y(x) := x\,\mathrm{I}\{x < y\}$, so that $W_1 = W$ and $T_1 = T$. Then $c\,W_y(X) = c\,y\,W(X/y)$ and $c\,T_y(X) = c\,y\,T(X/y)$. Now one can use the results of Subsections 2.1 and 2.2 with c, X, and σ replaced by $c\,y$, X/y, and σ/y, respectively. It should therefore be clear that the "cut" level was set to be 1 just for the simplicity of presentation.

Observe that for each $c > 0$ the exact lower bound $L_{W;c,\sigma}$ in (2.9) is no greater than 1, since the zero r.v. X obviously satisfies the conditions $\mathsf{E}X \geqslant 0$ and $\mathsf{E}X^2 \leqslant \sigma^2$. Hence, the exact lower bounds $L_{W;\sigma}$ and $L_{T;c,\sigma}$, which are no greater than $L_{W;c,\sigma}$, are as well no greater than 1. It is also clear that each of these exact lower bounds is nondecreasing in σ – since the exactness is over all r.v.'s X with $\mathsf{E}X \geqslant 0$ and $\mathsf{E}X^2 \leqslant \sigma^2$.

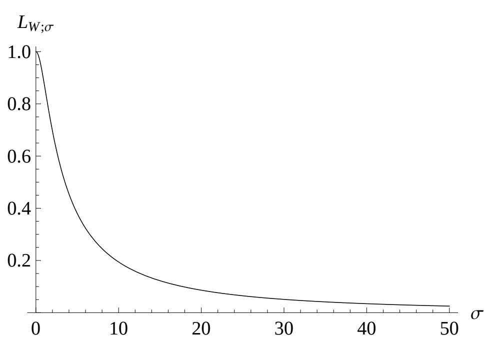

FIG 1. *The exact lower bound $L_{W;\sigma}$.*

However, for any $c > 0$ the exact lower bound $L_{W;c,\sigma}$ for the Winsorized r.v. $W(X)$ decreases rather slowly from 1 to 0 as σ increases from 0 to ∞. Even the smaller, universal over all $c > 0$ exact lower bound $L_{W;\sigma}$ decreases rather slowly; see Figure 1 and also recall Proposition 2.3. In particular, for the value $\sigma^2 = 1$ (which is of special interest as far as the application in [16] is concerned) the lower bound $L_{W;\sigma}$ is $0.878\ldots$, rather close to 1. Even for $\sigma^2 = 100$, this bound is $0.194\ldots$, not very small.

Moreover, the universal (over all $c > 0$) Winsorization bound $L_{W;\sigma}$ is remarkably close to the fixed-c Winsorization bounds $L_{W;c,\sigma}$, especially if the value of c is in the interval $[1,3] = \{\frac{p}{2}\colon 2 \leqslant p \leqslant 6\}$ – which is of particular interest in [16]. See the picture at the top of Figure 2; the green graph there, for $c = 2$ – cf. (2.19) – looks exactly horizontal at level 1, but it is in fact not.

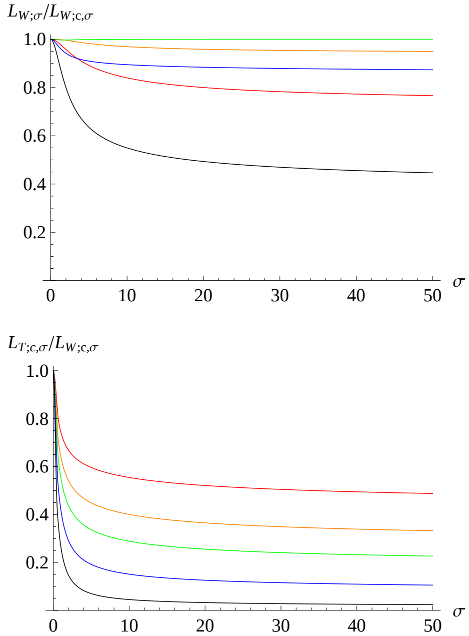

FIG 2. *The ratios $L_{W;\sigma}/L_{W;c,\sigma}$ (above) and $L_{T;c,\sigma}/L_{W;c,\sigma}$ (below) for $c = 1, \frac{3}{2}, 2, 3, 5$ – red, orange, green, blue, black, respectively.*

As for the truncation case, it is quite different from the Winsorization one. Indeed, the exact lower bound $L_{T;c,\sigma}$ is significantly smaller than $L_{W;c,\sigma}$, especially for larger values of σ. The bottom picture of Figure 2 shows the graphs of the ratios of these two bounds for $c = 1, \frac{3}{2}, 2, 3, 5$.

It is also easy to compare the asymptotics for $L_{T;c,\sigma}$ in (2.25) and (2.26) with that for $L_{W;c,\sigma}$ in (2.12) and (2.13). Comparing (2.12) with (2.25), it is easy to see that for $\sigma \downarrow 0$ one has $1 - L_{T;c,\sigma}$ is at least 4 times as large (asymptotically) as $1 - L_{W;c,\sigma}$, and may be infinitely many times as large when c goes to 0 or ∞. Similarly, for $\sigma \to \infty$, $L_{W;c,\sigma}$ is e^c times as large (asymptotically) as $L_{T;c,\sigma}$.

Moreover, in contrast with the Winsorization case, there is no nontrivial lower bound in the truncation case that would be universal over all $c > 0$. Namely, for any given $\sigma > 0$, the infimum of $\mathsf{E}\exp\{c\,T(X)\}$ over all $c > 0$ and all r.v.'s X with $\mathsf{E}X \geqslant 0$ and $\mathsf{E}X^2 \leqslant \sigma^2$ is 0; the same holds even if the conditions $\mathsf{E}X \geqslant 0$ and $\mathsf{E}X^2 \leqslant \sigma^2$ are strengthened to $\mathsf{E}X = 0$ and $\mathsf{E}X^2 = \sigma^2$. Indeed, let $a \downarrow 0$, $b := \sigma^2/a$, and $c := 1/a^2$; then it is easy to see that $\mathsf{E}\exp\{c\,T(X_{a,b})\} \to 0$; cf. (2.25) and (2.26), with large c.

The general problem of finding the maximum or minimum of the generalized moment $\int f\,\mathrm{d}\mu$ over the set of all nonnegative measures with given generalized moments $\int f_i\,\mathrm{d}\mu$ $(i \in I)$ goes back to Chebyshev and Markov; here a function f and a family of functions $(f_i)_{i\in I}$ are given; see, e.g., [6, 8–11, 13]. One group of results in this area is that for finite I under general conditions it may be assumed without loss of generality that the support of μ is also finite, with cardinality no greater than that of I; methods based on such results may be referred to as finite-support methods. Other results, valid for finite or infinite I, concern the following duality: under general conditions, the supremum (say) of $\int f\,\mathrm{d}\mu$ over all μ such that $\int f_i\,\mathrm{d}\mu \leqslant c_i$ for all $i \in I$ coincides with the infimum of $\int_I c_i\,\nu(\mathrm{d}i)$ over all nonnegative measures ν on I (say with a finite support) such that $\int_I f_i\,\nu(\mathrm{d}i) \geqslant f$.

Such methods were used e.g. in [1–3, 7, 12, 17–19]. In particular, the supremum of $\int_{\mathbb{R}} e^x\,\mu(\mathrm{d}x)$ given $\int_y^\infty \mu(\mathrm{d}x) = 0$, $\int_{\mathbb{R}} x\,\mu(\mathrm{d}x) = 0$, and $\int_{\mathbb{R}} x^2\,\mu(\mathrm{d}x) = \sigma^2$ was found (implicitly) in [1] and (explicitly) in [17].

In [7] a similar problem was solved, under the additional restriction that μ is a

probability measure. This result was extended in [2, 3] to the Eaton-like moment functions $(\cdot - t)_+^2$ $(t \in \mathbb{R})$ in place of the exponential function $e^{\cdot}$ in [7]; on the other hand, this was a further development of the line of results obtained in [4, 5, 13, 14]. The results of [1, 7] and [2, 3] were refined in [17] and [12], respectively, by also taking into account positive-part third moments. The supremum of the moments $\int f\,\mathrm{d}\mu$ over all Stein-type moment functions $x \mapsto f(x) := xg(x) - g'(x)$ with Lipschitz-1 functions g' and over all probability measures μ with given mean, variance, and third absolute moment was presented (in an equivalent form) in [18, Theorem 3]. Results somewhat related to the mentioned ones were obtained in [15]; see also the bibliography therein. Of course, mentioned above are a very small sample of the work done on the Chebyshev-Markov type of extremal problems.

Concerning our problems of minimizing the exponential moments of $W(X)$ and $T(X)$, one could use mentioned finite-support methods to reduce the consideration to r.v.'s X taking only three values, since we have here three affine restrictions: on the first two moments and on the total mass of the measure (which has to be a probability measure). Another, more ad hoc kind of approach would be to condition the distribution of the r.v. X on $\mathrm{I}\{X < 1\}$, which would preserve the mean and would not increase the second moment; also, this conditioning would not increase the exponential moments of $W(X)$ and $T(X)$, since both functions, e^{cW} and e^{cT}, are convex on $(-\infty, 1)$ and on $[1, \infty)$; thus, it would remain to consider r.v.'s X taking only two values. However, the duality-type method that we chose to prove (in the next section) inequalities (2.9) and (2.24) appears more effective, as it immediately reduces the consideration to r.v.'s $X_{a,b}$ that, not only take just two values, but also have the first two moments exactly equal to 0 and σ^2, respectively; moreover, this approach appears more convenient in obtaining the strictness conditions for inequalities (2.9) and (2.24).

3. Proofs

Proof of Proposition 2.1.

(I). Part (I) follows because $ab^*_{a,c}$ strictly and continuously increases from 0 to ∞ as a increases from 0 to ∞.

(II). Observe that $a(a^2+\sigma^2)^2\ell_1'(a)$ is a quadratic polynomial in σ^2, whence one can see that the system of inequalities $\ell_1'(a) > 0$ and $0 < a < \sigma^2$ can be rewritten as $0 < a < \sqrt{1+\sigma^2} - 1$. This means that $\ell_1(a)$ switches from increase to decrease over $a \in (0, \sigma^2)$; at that, $\ell_1(0+) = -\infty$ and $\ell_1(\sigma^2) = 0$. Now part (II) of Proposition 2.1 follows as well. $\square$

Proof of Theorem 2.2. Let

$$F(x) := e^{c\,W(x)} \quad \text{and} \quad G(x) := \alpha + \beta x + \gamma x^2 \tag{3.1}$$

for all $x \in \mathbb{R}$, where

$$\alpha := e^c - \frac{b^2 c e^{-ac}}{2(a+b)}, \quad \beta := \frac{bce^{-ac}}{a+b}, \quad \gamma := -\frac{ce^{-ac}}{2(a+b)}, \tag{3.2}$$

$a > 0$, $b > 1$, $c > 0$. Then it is straightforward to check that $F(b) = G(b)$, $F'(b) = G'(b)$, and $F'(-a) = G'(-a)$. Let now $b = b^*_{a,c}$. Then $b > 2 > 1$ and $F(-a) = G(-a)$, so that

$$F(b) = G(b), \quad F'(b) = G'(b), \quad F(-a) = G(-a), \quad F'(-a) = G'(-a). \tag{3.3}$$

Also, by (3.2), $\gamma < 0$ and hence the function G is strictly concave, while the function F is convex on $(-\infty, 1)$ and on $[1, \infty)$; so, the difference $D := F - G$ is strictly convex on $(-\infty, 1)$ and on $[1, \infty)$; at that, $D(-a) = D'(-a) = D(b) = D'(b) = 0$, whence $D > 0$ and $F > G$ on $\mathbb{R} \setminus \{-a, b\}$, while $F = G$ on the two-point set $\{-a, b\}$. Now specify a, to $a = a_{c,\sigma}$. Then, recalling (2.6) and (2.3), one sees that

$$b^*_{a_{c,\sigma},c} = b_{c,\sigma}. \tag{3.4}$$

Therefore, also specifying b to $b = b_{c,\sigma}$, one has

$$\mathsf{E} e^{c\,W(X)} = \mathsf{E} F(X) \geqslant \mathsf{E} G(X) \geqslant \mathsf{E} G(X_{a,b}) = \mathsf{E} F(X_{a,b}) = \mathsf{E} \exp\big\{c\,W(X_{a,b})\big\}; \tag{3.5}$$

the second inequality here takes place because (in view of (3.2)) $\beta > 0 > \gamma$, while $\mathsf{E}X \geqslant 0 = \mathsf{E}X_{a,b}$ and $\mathsf{E}X^2 \leqslant \sigma^2 = ab = \mathsf{E}X_{a,b}$. Thus, (2.9) follows. Moreover, because $F > G$ on $\mathbb{R} \setminus \{-a, b\}$, the first inequality in (3.5) is strict unless the support of the distribution of X is a subset of $\{-a, b\}$, and the second inequality in (3.5) is strict unless $\mathsf{E}X = 0$ and $\mathsf{E}X^2 = \sigma^2 = ab$; thus indeed, inequality (2.9) is strict unless $X \stackrel{\mathrm{D}}{=} X_{a,b}$.

To prove inequality (2.10), it suffices to show that $\mathsf{E}\exp\big\{c_\sigma\,W(X_{a_\sigma,b_\sigma})\big\}$ is a lower bound on $\mathsf{E}\exp\big\{c\,W(X_{a,b})\big\}$ for any a and b such that $a \in (0, \sigma^2)$ and $b = \sigma^2/a$. Take indeed any such a and b. Observe that $\mathsf{E}\exp\big\{c\,W(X_{a,b})\big\} = \frac{a^2e^c + e^{-ac}\sigma^2}{a^2+\sigma^2}$ is strictly convex in $c \in \mathbb{R}$ and attains its minimum,

$$m(a,\sigma) := \frac{a(1+a)\left(\frac{a}{\sigma^2}\right)^{-\frac{1}{1+a}}}{a^2+\sigma^2},$$

in c only at $c = \frac{\ln(\sigma^2/a)}{1+a} = \frac{\ln b}{1+a}$ – cf. (2.8). Next, the minimum of $m(a,\sigma)$ or, equivalently, of

$$\ell(a) := \ell(a,\sigma) := \ln m(a,\sigma)$$

in $a \in (0, \sigma^2)$ is attained only at the point $a = a_\sigma$ defined by (2.5), because, by part (II) of Proposition 2.1, $\ell'(a) = \ell_1(a)/(1+a)^2$ switches in sign from $-$ to $+$ over $a \in (0, \sigma^2)$. Thus, inequality (2.10) is true, and it is strict unless $c = c_\sigma$ and $a_{c,\sigma} = a_\sigma$.

It remains to verify the three equalities in (2.11). In view of (3.4), the last of these equalities is implied by the first one. So, if any of the equalities in (2.11) were false, then the equality $X_{a_\sigma,b_\sigma} \stackrel{\mathrm{D}}{=} X_{a_{c_\sigma,\sigma},b_{c_\sigma,\sigma}}$ would also be false, and so, by what has been proved, inequality (2.9) with c_σ and X_{a_σ,b_σ} in place of c and X would be strict, which would contradict inequality (2.10).

This completes the proof of Theorem 2.2. $\square$

Proof of Proposition 2.3.

(I) To prove part (I), consider first the case $\sigma \downarrow 0$. Then, by (2.3) and (2.2), $a_{c,\sigma} \downarrow 0$. Moreover, $b^*_{a,c} \to \frac{2(e^c-1)}{c}$ whenever $a \to 0$. So, by (3.4), $b_{c,\sigma} \to \frac{2(e^c-1)}{c} > 2 > 1$, and so, $a_{c,\sigma} = \sigma^2/b_{c,\sigma} \sim \frac{c}{2(e^c-1)}\,\sigma^2$. On the other hand,

$$\mathsf{E}\exp\{c\,W(X_{a,b})\} - 1 = (e^c-1)\,\tfrac{a}{a+b} + (e^{-ca}-1)\,\tfrac{b}{a+b} \sim (\tfrac{e^c-1}{b} - c)a \tag{3.6}$$

whenever $a \downarrow 0$ and $a = o(b)$. This, together with the relations $b_{c,\sigma} \to \frac{2(e^c-1)}{c}$ and $a_{c,\sigma} \sim \frac{c}{2(e^c-1)}\,\sigma^2$, implies (2.12).

Consider now the case $\sigma \to \infty$. Then, by (2.3) and (2.2), $a_{c,\sigma} \to \infty$. Next,

$$b^*_{a,c} \sim \tfrac{2e^c}{c}\,e^{ac} \quad \text{as } a \to \infty. \tag{3.7}$$

So, in view of (2.6) and (3.4), for $a = a_{c,\sigma}$ one has

$$\sigma^2 \sim \tfrac{2e^c}{c^2}\,ac\,e^{ac} = e^{ac\,(1+o(1))}, \tag{3.8}$$

whence $a_{c,\sigma} = a \sim \frac{1}{c}\ln(\sigma^2)$ and $b_{c,\sigma} = \sigma^2/a_{c,\sigma} \sim c\sigma^2/\ln(\sigma^2)$. Also, for $a \to \infty$ and $b = b^*_{a,c}$ (3.7) yields $a = o(b)$ and

$$\mathsf{E}\exp\{c\,W(X_{a,b})\} = \tfrac{ae^c + be^{-ac}}{a+b} \sim \tfrac{ae^c + 2e^c/c}{a+b} \sim \tfrac{ae^c}{b} \tag{3.9}$$

This, together with the relations $a_{c,\sigma} \sim \frac{1}{c}\ln(\sigma^2)$ and $b_{c,\sigma} \sim c\sigma^2/\ln(\sigma^2)$, implies (2.13).

(II) Note that $f(0+) = -\infty$, $f(1) = 0$, and $f'(t) = -2(t-\frac{1}{2})/t$ switches in sign from $+$ to $-$ as t increases from 0 to 1. Now part (II) of Proposition 2.3 follows.

(III) To prove part (III), consider first the case $\sigma \downarrow 0$. Then, by (2.4) and (2.14), for each fixed $t \in (0,1)$ one has $\ell_1(t\sigma^2) = \ln t - (2+o(1))(t-1) = f(t) + o(1)$. So, by part (II) of Proposition 2.3 for each fixed $t \in (0,t_*)$ one has $\ell_1(t\sigma^2) < 0$ – eventually, for all small enough σ; similarly, for each fixed $t \in (t_*,1)$ eventually $\ell_1(t\sigma^2) > 0$. Therefore, by (2.5), $a_\sigma \sim t_*\sigma^2$ and hence, by (2.7) and (2.8), $b_\sigma \to 1/t_*$ and $c_\sigma \to -\ln t_*$. Now (2.16) follows by (3.6), since for $c = c_\sigma$ one has $e^c \to 1/t_*$ and $c \to -\ln t_* = 2(1-t_*)$, the last equality due to (2.15)-(2.14).

The case $\sigma \to \infty$ is considered similarly. Then, by (2.4), $\ell_1\big(\kappa\ln(\sigma^2)\big) \sim 2(\kappa-\frac{1}{2})\ln(\sigma^2)$ for each fixed $\kappa \in (0,\infty)\setminus\{\frac{1}{2}\}$, so that $\ell_1\big(\kappa\ln(\sigma^2)\big)$ is eventually less than 0 for each $\kappa \in (0,\frac{1}{2})$ and eventually greater than 0 for each $\kappa \in (\frac{1}{2},\infty)$. Thus, by part (II) of Proposition 2.1, $a_\sigma \sim \frac{1}{2}\ln(\sigma^2)$ and hence $b_\sigma \sim 2\sigma^2/\ln(\sigma^2)$ and $c_\sigma \to 2$. Moreover, by (2.11), one has $b_\sigma = b^*_{a_\sigma,c_\sigma}$. Recall that relations (3.7) and (3.9) were derived assuming that $a \to \infty$, $b = b^*_{a,c}$, and $c > 0$ is fixed. Reasoning quite similarly – with a_σ, b_σ, $c_s i$ in place of such a, b, c – one concludes that $L_{W;\sigma} \sim a_\sigma\,e^{c_\sigma}/b_\sigma$, and now (2.17) follows since $a_\sigma \sim \frac{1}{2}\ln(\sigma^2)$, $b_\sigma \sim 2\sigma^2/\ln(\sigma^2)$, and $c_\sigma \to 2$.

(IV) The derivative of $\frac{-c^2}{4(e^c-1)}$ in $c > 0$ is positive iff $f(t) < 0$ for $t := e^{-c}$. Therefore and by part (II) of Proposition 2.3, $\frac{-c^2}{4(e^c-1)}$ attains a minimum in

$c > 0$ at $c = -\ln t_* = 2(1 - t_*)$. Replacing now c in the denominator of $\frac{-c^2}{4(e^c-1)}$ by $-\ln t_*$ and in the numerator, by $2(1 - t_*)$, one obtains (2.18). As for (2.19), it is much more straightforward. $\square$

Proof of Proposition 2.4. That each of the equations on the right-hand sides of (2.21) and (2.22) has a unique root $a > 0$ follows because both $B^*_{a,c}$ and $aB^*_{a,c}$ strictly and continuously increase from 0 to ∞ as does so. Now (2.23) follows because the value of $a\,B^*_{a,c}$ at $a = A_c$ is A_c. $\square$

Proof of Theorem 2.5. Let here $F(x) := e^{c\,T(x)}$ and let $G(x)$ be defined as in (3.1).

Consider first the case $\sigma^2 \leqslant A_c$. Here, take $G(x)$ with

$$\alpha := \frac{e^{-ac}(a^2 e^{ac} + ca^2 + ac + 2a + 1)}{(a+1)^2},$$
$$\beta := \frac{e^{-ac}\big(2a(e^{ac} - 1) + c(1 - a^2)\big)}{(a+1)^2}, \quad \gamma := \frac{e^{-ac}(e^{ac} - ac - c - 1)}{(a+1)^2}.$$

Then for $D := F - G$ and any $a > 0$ one has $D(-a) = D'(-a) = D(1) = 0$. Let now $a = \sigma^2$, so that (by the current case condition) $0 < a \leqslant A_c$, which implies $B^*_{a,c} \leqslant 1$ and hence $G'(1) \leqslant 0$ $\big($because $G'(1) = (B^*_{a,c} - 1)c\,e^{-ac}/(1 + a)\big)$; also, $0 < a \leqslant A_c$ implies $(e^{ac} - 1) - (1 + a)c < 2(e^{ac} - 1) - (1 + a)c = (B^*_{a,c} - 1)c \leqslant 0$, whence $\gamma < 0$, so that G is strictly convex on $\mathbb{R}$; moreover, $\beta > e^{-ac}c\,(1 + a^2)/(1 + a)^2 > 0$. In turn, the inequality $G'(1) \leqslant 0$ means that $D'(1+) \geqslant 0$; also, D is strictly convex on $(-\infty, 1)$ and $[1, \infty)$; recalling now that $D(-a) = D'(-a) = D(1) = 0$, one has $D > 0$ and $F > G$ on $\mathbb{R} \setminus \{-a, 1\}$, while $F = G$ on the two-point set $\{-a, 1\} = \{-\sigma^2, 1\}$. Now the first line of (2.24) follows – cf. (3.5).

Consider now the case $\sigma^2 \geqslant A_c$. Here, take $G(x)$ with

$$\alpha := \frac{e^{-ac}(ca^2 + 2abc + 2a + 2b)}{2(a+b)}, \quad \beta := \frac{cbe^{-ac}}{a+b}, \quad \gamma := -\frac{ce^{-ac}}{2(a+b)},$$

where $a > 0$ and $b := B^*_{a,c}$. Assume now also that a is so large as $b \geqslant 1$. Then, again for $D := F - G$, one has $D(-a) = D'(-a) = D(b) = D'(b) = 0$, while $\beta > 0 > \gamma$, so that again D is strictly convex on $(-\infty, 1)$ and $[1, \infty)$, $D > 0$ and $F > G$ on $\mathbb{R} \setminus \{-a, b\}$, while $F = G$ on the two-point set $\{-a, b\}$; if $b = 1$ then $D'(b)$ is understood as the right derivative of D at point 1. Since the current case if $\sigma^2 \geqslant A_c$, (2.23) yields $A_{c,\sigma} \geqslant A_c$ and hence $B_{c,\sigma} = \sigma^2/A_{c,\sigma} = B^*_{A_{c,\sigma},c} \geqslant B^*_{A_c,c} = 1$. Now, reasoning again similarly to (3.5), one obtains the second line of (2.24).

The proof of the strictness statement on (2.24) is quite similar to that for (2.9), because here as well one has $\beta > 0 > \gamma$ – in either case, whether $\sigma^2 \leqslant A_c$ or $\sigma^2 \geqslant A_c$. $\square$

Proof of Proposition 2.6.

Here the case $\sigma \downarrow 0$ is quite staightforward. Indeed, then, by (2.24), one eventually has $L_{T;c,\sigma} - 1 = \mathsf{E}\exp\{c\,T\big(X_{\sigma^2,1}\big)\} - 1 = \frac{e^{-c\,\sigma^2}-1}{1+\sigma^2} \sim -c\,\sigma^2$.

As for the case $\sigma \to \infty$, the proof of (2.26) is quite similar to that of relation (2.13) in part (I) of Proposition 2.3: replace all instances of $b^*_{a,c}$, $a_{c,\sigma}$, $b_{c,\sigma}$, W by $B^*_{a,c}$, $A_{c,\sigma}$, $B_{c,\sigma}$, T, respectively, and also drop all instances of the factor e^c in the numerators of the ratios in (3.7), (3.8), and (3.9). $\square$

References

[1] Bennett, G. Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc. 57*, 297 (1962), 33–45.

[2] Bentkus, V. A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand. *Liet. Mat. Rink. 42*, 3 (2002), 332–342.

[3] Bentkus, V. On Hoeffding's inequalities. *Ann. Probab. 32*, 2 (2004), 1650–1673.

[4] Eaton, M. L. A note on symmetric Bernoulli random variables. *Ann. Math. Statist. 41* (1970), 1223–1226.

[5] Eaton, M. L. A probability inequality for linear combinations of bounded random variables. *Ann. Statist. 2* (1974), 609–613.

[6] Hoeffding, W. The extrema of the expected value of a function of independent random variables. *Ann. Math. Statist. 26* (1955), 268–275.

[7] Hoeffding, W. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc. 58* (1963), 13–30.

[8] Karlin, S., and Studden, W. J. *Tchebycheff systems: With applications in analysis and statistics.* Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.

[9] Karr, A. F. Extreme points of certain sets of probability measures, with applications. *Math. Oper. Res. 8*, 1 (1983), 74–85.

[10] Kemperman, J. H. B. On the role of duality in the theory of moments. In *Semi-infinite programming and applications (Austin, Tex., 1981)*, vol. 215 of *Lecture Notes in Econom. and Math. Systems.* Springer, Berlin, 1983, pp. 63–92.

[11] Kreĭn, M. G., and Nudel′man, A. A. *The Markov moment problem and extremal problems.* American Mathematical Society, Providence, R.I., 1977. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50.

[12] Pinelis, I. On the Bennett-Hoeffding inequality (preprint), arxiv: 0902.4058v1 [math.PR].

[13] Pinelis, I. Optimal tail comparison based on comparison of moments. In *High dimensional probability (Oberwolfach, 1996)*, vol. 43 of *Progr. Probab.* Birkhäuser, Basel, 1998, pp. 297–314.

[14] Pinelis, I. Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. In *Advances in stochastic*

inequalities (Atlanta, GA, 1997), vol. 234 of *Contemp. Math.* Amer. Math. Soc., Providence, RI, 1999, pp. 149–168.

[15] Pinelis, I. Optimal two-value zero-mean disintegration of zero-mean random variables. *Electron. J. Probab. 14* (2009), no. 26, 663–727.

[16] Pinelis, I., and Molzon, R. Berry-Esseen bounds for general nonlinear statistics, with applications to Pearson's and non-central Student's and Hotelling's (preprint, arxiv:0906.0177v1 [math.ST]).

[17] Pinelis, I. S., and Utev, S. A. Sharp exponential estimates for sums of independent random variables. *Theory Probab. Appl. 34*, 2 (1989), 340–346.

[18] Tyurin, I. New estimates of the convergence rate in the Lyapunov theorem (preprint, arxiv:0912.0726v1 [math.PR]).

[19] Utev, S. A. Extremal problems in moment inequalities. In *Limit theorems of probability theory*, vol. 5 of *Trudy Inst. Mat.* "Nauka" Sibirsk. Otdel., Novosibirsk, 1985, pp. 56–75, 175.